APPENDIX - A

Hamilton Discharge Data

Appendix A: Hamilton Discharge Data

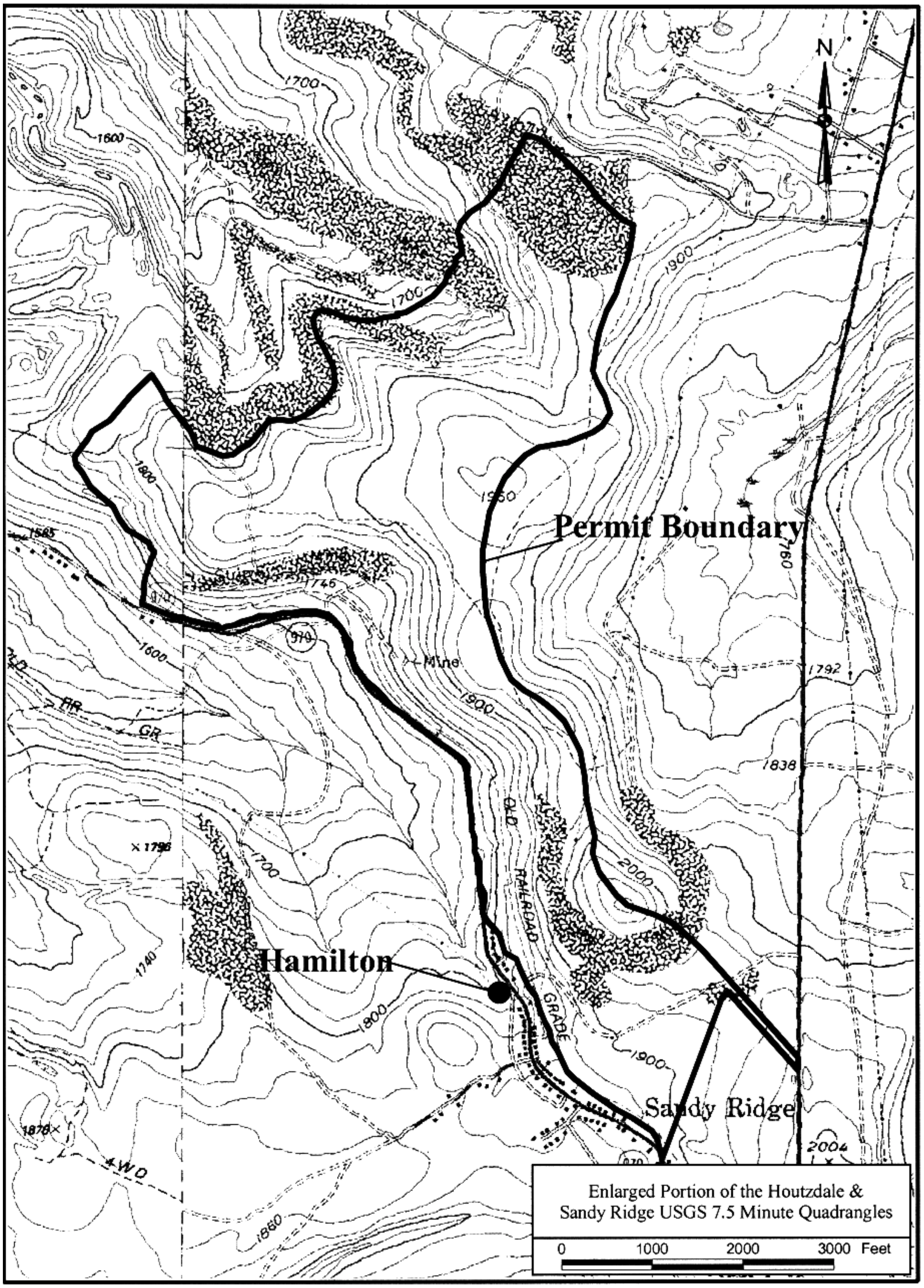
The Hamilton site is a permitted remining site located in Clearfield County, Pennsylvania, as shown in Figure A.1. Several years of background (pre-mining) baseline data existed for two abandoned mine discharges on the site (Hamilton 01 and Hamilton 08). This site was selected to be the initial data set statistically analyzed by Dr. J.C. Griffiths during February to April of 1987. The first two reports of the eight report series of statistical analyses completed by Dr. Griffiths in 1987 and 1988 were on the Hamilton site. Report No. 1 was a preliminary evaluation of the MINITAB[1] software package for the analysis of remining data, performed on the Hamilton 01 and 08 data files. Report No. 2 was an evaluation of the usefulness of MINITAB in conducting a time series analysis, including Box-Jenkins procedures, of the Hamilton 08 discharge data set. Since these first two reports were preliminary or exploratory in nature, they were not as well developed as far as evaluation of the various steps of the data analysis algorithm (see Figure 3.1) as succeeding reports (Report Nos. 3 – 8). These succeeding reports are the subject of Chapters 4 through 8 of this report (Report No. 8 of the original Griffiths report was a synopsis of Report Nos. 1 to 7). However some items of interest, not found in the other reports, were expressed in the Hamilton site reports, and the data set is a good example of remining permit data. Thus, it was determined that the elements of these two reports (although somewhat sketchy in places) and the data sets would be presented in this Appendix.

The Hamilton 01 data had problems (missing data and the presence of a few exceptionally high values) similar to the other data sets described in Chapters 4 through 8. For high values of manganese and sulfate, for example, it was stated that it is important to decide whether to keep or reject the values as outliers with the assumption they are data recording errors and therefore, not really meaningful. Examination of each example, case by case, is recommended to make an appropriate decision. Logarithmic transformation of some variables was attempted, but introduced negative skewness in the sulfate data. Ultimately, it was determined that the sulfate data appeared to be acceptable without transformation.

Some univariate and bivariate analyses were conducted on the Hamilton 08 discharge data. It was found that there was no obvious relationship between flow and acidity. There also was no apparent relationship between acidity and sulfate. There seems to be a weak inverse relationship between manganese and flow (flow increases as manganese decreases). Simple time series plots of acidity, iron, manganese, and sulfate data from the Hamilton 08 discharge were also performed, and some obvious cycles were observed.

[1]MINITAB is a commercial software package from Minitab, Inc. ©1986, 3081 Enterprise Drive, State College, PA 16801.

Figure A.1: Map of Hamilton Site



Following application of stem and leaf plots, box plots, scatterplots and time series plots, it was determined that Cross-Correlation functions, Rootogram functions, and the Box-Jenkins procedures in the software package should be applied. It was also concluded that many additional analytical tools could be used including analysis of variance, t-tests, Chi-Square tests, and regression. It is necessary to emphasize that while these tests are easy to apply, both their applicability and the interpretation of the results may be very demanding.

The objective of Griffiths Report No. 2 was an attempt to fit a model to the Hamilton 08 discharge data, and preferably, to find a single simple model that would provide a reasonably close fit. It is desirable to find a single model, if feasible, for all five variables. The Box-Jenkins time series analysis procedure was used for this purpose (Box and Jenkins, 1970). This procedure consists of a convenient package of computer programs that embrace the entire modeling process. A wide variety of models, collectively known as the ARIMA models, is available in this package. Use of this sophisticated procedure requires that the data be collected at equal time intervals. This requirement was only partially fulfilled by the Hamilton 08 data. Therefore, application of the resulting model(s) should be limited.

Eventually, when the model meets the demands of the criteria, it may be used to forecast future values of the variable, accompanied by an appropriate estimate of the confidence limits at a selected probability level. Any new observations may be added to the chosen model and the fit examined for acceptance or rejection. These data should be taken at the same time intervals as the original series (i.e., if the original observations are taken at two week intervals the new observations should also be taken at two week intervals). The number of samples need not be extensive; six to twelve would be acceptable.

If second differences of the flow data set are taken, the Acf and Pacf show many large spikes suggesting that the series has been overdifferenced. It therefore seems evident that an MA (0,1,1) model may be most suitable. The first check criterion is a measure of correlation among the parameters. Since, in this case, there is only one parameter, this does not apply. The second criterion is the Acf of the residuals; if the model "fits" well, all systematic variation has been removed and the remainder is random (equals white noise). There are two tests at this stage: the first is an overall Portmanteau test (Box-Pierce-Ljung Statistic) of all autocorrelations taken together. For this case the result is $X^2 = 17.95$ with 29 degrees of freedom. It is not significantly greater than that expected from white noise, hence it is feasible to consider that these residuals represent random variation and, on the basis of this criterion, there is no evidence to reject the model.

The second test is to examine the individual autocorrelations against twice their standard errors. Since none exceed this value there is no evidence to require further refinement in the model. When first differences of the residuals are taken, the Portmanteau test yields a highly significant value, implying overdifferencing. The Pacf of the residuals confirms this diagnosis.

A number of alternative models were fitted to log-transformed flow data and the results are summarized in Table A.1. Both the AR (1,0,0) and MA (0,1,1) models fit equally well. The coefficient (Φ) in the AR model is approximately equivalent to the first difference in the MA model. Attempts to improve on these simple models by using additional coefficients, seasonal and otherwise, failed to provide any substantial improvement. Thus, it was decided to select one of the simpler models.

Table A.1: Alternate Models Fitted to Log Flow Data

No.	Model	Residual Sum of Squares	Coefficients	Acf spikes	Portmanteau Chi-Square statistic	Residual Standard Deviation
1	MA (0,1,1)	30.73	*	None	17.95	0.519
2	AR (1,0,0)	28.57	*	None	27.58	0.503
3	AR (1,1,1)	30.34	θ_1 not significantly different from 0	None	17.89	0.520
4	AR (2,0,0)	28.57	Φ_1, Φ_2 significantly correlated	None	27.58	0.505
5	AR (1,1,0)	30.83	*	None	19.88	0.522
6	AR (1,0,0) (1,0,0)	25.72	*	None	22.11	0.481
7	AR (1,0,0) (0,0,1)	29.81	Φ_1 not significantly different from 0	None	13.40	0.523

* All coefficients are significantly different from 0 or 1, and there are not significant correlations between coefficients

It was concluded that the most appropriate model, common to all variables, is the simple moving average of the first differences of the observations, or an MA (0,1,1) model. The resulting equations for each variable are:

Log Flow $z_t = z_{t-1} + a_t - 0.415a_{t-1}$

Log Acidity $z_t = z_{t-1} + a_t - 0.381a_{t-1}$

Log Fe $z_t = z_{t-1} + a_t - 0.824a_{t-1}$

Log Mn $z_t = z_{t-1} + a_t - 0.662a_{t-1}$

$S0_4$ $z_t = z_{t-1} + a_t - 0.408a_{t-1}$

The model implies that the observation at time t (z_t) equals its previous value plus a contribution from the shock term (a_t) and an additional, smaller contribution from the shock term of the previous period (a_{t-1}). The system appears to have only a one-step memory and is otherwise a typical random variable.

The absence of a seasonal component may be attributed to the fact that there are extreme variations in the data which tend to smother any smaller systematic contribution. There appears to be two main reasons for this, one of which may be modified. The first is the presence of zeros in the data and the absence of an attempt to smooth the data. Smoothing may well be of major importance in reducing the effects of extreme variations and thus, reducing the confidence limits around forecasts. The second reason is that the unusual events represented by large positive residuals are not repeated at the same interval during each annual cycle. Thus, a heavy influx of water from spring melt is common but is not consistently heavy, and rarely occurs on the same date. Again, there are heavy late spring storms which lead to flooding, but do not occur every year and do not always occur in the same month. Thus the spread of events from February to June would tend to smooth out any persistent cyclical feature that may be present. A much longer series would be needed to check these possible effects.

There is one other aspect to the data that may be of importance. It may not be desirable to perform a test of the observations that is too stringent, because it could result in too many false alarms. Thus, a fairly simple, robust test is desirable in practice. The present MA models may well be adequate for this purpose. Investigations at more locations may help to clarify these questions.

Hamilton 01

Rows	Days	Flow	Acidity	Total Iron	Manganese	Sulfate
1	0	9.70	337.0	43.30	5.12	868.0
2	34	3.00	360.0	20.00	6.80	900.0
3	73	10.00	305.0	14.00	9.00	652.0
4	119	0.00	400.0	33.00	8.00	823.0
5	161	0.00	294.0	25.00	3.00	422.0
6	202	0.00	307.0	19.00	4.00	550.0
7	216	0.00	305.0	27.00	4.00	342.0
8	231	0.00	300.0	19.00	4.10	419.0
9	244	28.61	539.0	8.50	3.00	142.0
10	257	7.71	195.0	35.00	2.90	430.0
11	271	28.00	174.0	16.20	5.00	494.0
12	286	7.70	184.0	16.00	8.20	510.0
13	299	7.70	230.0	21.50	4.00	600.0
14	313	7.70	306.0	28.00	7.00	612.0
15	329	7.70	254.0	28.00	11.30	382.0
16	342	7.70	394.0	35.00	10.00	423.0
17	356	7.70	444.0	19.00	5.30	705.0
18	369	7.70	340.0	35.00	7.20	399.0
19	383	0.00	474.0	75.00	8.00	872.0
20	386	2.10	714.0	75.00	8.20	608.0
21	411	82.00	222.0	62.00	4.60	550.0
22	425	7.70	258.0	54.30	7.70	500.0
23	455	28.60	274.0	7.00	7.00	550.0
24	467	28.00	282.0	42.00	56.00	700.0
25	482	28.00	284.0	20.00	5.10	510.0
26	495	28.00	268.0	20.00	5.90	681.0
27	510	61.30	220.0	9.50	4.80	620.0
28	524	105.00	202.0	7.20	4.90	598.0
29	538	28.00	214.0	10.00	5.20	613.0
30	552	105.00	110.0	3.70	2.30	587.0
31	565	105.00	118.0	8.00	2.90	358.0
32	579	105.00	162.0	14.70	5.00	469.0
33	593	28.00	224.0	25.00	3.80	655.0
34	608	28.00	250.0	25.00	4.90	713.0
35	624	28.00	98.0	12.50	4.50	477.0
36	636	105.00	197.0	19.00	4.00	397.0
37	650	28.00	78.0	6.00	3.70	612.0
38	666	7.70	264.0	18.00	5.00	600.0
39	680	28.00	218.0	5.20	3.00	542.0
40	692	28.00	286.0	10.00	5.90	643.0
41	706	0.00	458.0	13.00	6.50	746.0
42	721	7.70	352.0	8.50	7.50	811.0
43	734	0.00	356.0	6.90	7.30	778.0
44	748	2.10	632.0	9.30	7.90	568.0
45	762	7.70	392.0	8.70	9.30	831.0
46	772	2.10	364.0	9.00	8.00	806.0
47	790	7.70	336.0	7.50	8.00	835.0

Rows	Days	Flow	Acidity	Total Iron	Manganese	Sulfate
48	799	7.70	334.0	7.00	8.00	798.0
49	818	7.70	316.0	9.00	7.00	655.0
50	832	7.70	306.0	9.00	6.00	713.0
51	846	7.70	370.0	12.80	4.90	794.0
52	857	7.70	306.0	9.00	8.00	674.0
53	867	7.70	346.0	9.00	8.00	719.0
54	874	28.00	206.0	8.00	5.53	431.0
55	885	28.00	264.0	8.00	6.00	566.0
56	899	28.00	264.0	10.00	6.00	594.0
57	916	105.00	261.0	3.75	4.00	333.0
58	930	61.00	226.8	2.25	3.00	29.6
59	944	61.00	155.6	4.00	3.00	276.0
60	958	61.00	133.4	2.75	2.30	181.0
61	972	61.00	168.1	4.20	3.90	300.0
62	989	29.00	74.2	9.80	5.10	343.0
63	1000	28.97	142.2	8.00	7.80	491.0
64	1015	7.80	158.3	12.00	6.00	550.0
65	1028	0.00	191.3	7.50	9.00	511.0
66	1043	7.80	232.6	9.00	10.00	584.0
67	1052	7.80	266.8	7.50	7.00	690.0
68	1070	7.80	300.7	9.50	19.00	531.0
69	1085	7.80	317.3	8.00	14.00	452.0
70	1098	7.80	326.6	9.50	11.50	755.0
71	1116	1.20	314.7	6.00	10.00	805.0
72	1126	2.20	287.3	8.50	9.00	816.0
73	1141	7.90	265.6	7.50	9.00	780.0
74	1154	6.10	184.5	8.00	9.50	608.0
75	1171	8.90	121.2	9.00	4.40	300.0
76	1184	41.70	91.6	3.10	3.30	261.0
77	1197	2.20	166.1	5.50	4.80	396.0
78	1210	197.00	197.0	8.50	5.30	524.0
79	1221	0.00	226.1	6.50	7.30	652.0
80	1238	61.00	215.7	8.00	7.50	609.0
81	1248	131.00	84.6	1.20	2.90	187.0
82	1266	0.00	107.8	4.00	2.20	242.0
83	1280	7.90	126.7	6.50	3.50	337.0
84	1294	18.80	107.1	3.90	2.80	246.0
85	1308	12.10	128.2	6.00	2.80	264.0
86	1322	11.00	126.7	7.30	3.60	284.0
87	1336	8.90	124.7	6.30	2.90	255.0
88	1351	12.00	102.4	5.50	2.40	236.0
89	1365	2.70	190.4	9.00	6.30	455.0
90	1379	4.60	179.4	6.50	5.50	385.0
91	1393	4.60	189.3	7.00	4.30	481.0
92	1407	2.70	202.3	8.00	7.80	596.0
93	1421	4.60	664.7	8.00	6.50	466.0
94	1434	12.10	163.8	3.90	3.20	299.0
95	1450	4.60	194.3	3.90	4.60	466.0
96	1464	9.90	231.5	8.50	5.50	612.0

Rows	Days	Flow	Acidity	Total Iron	Manganese	Sulfate
97	1477	5.40	265.6	9.50	9.80	700.0
98	1487	2.70	264.7	8.50	9.80	1223.0
99	1504	4.60	447.7	1.60	10.60	716.0
100	1515	4.60	287.2	1.82	10.30	667.0
101	1526	2.20	282.8	11.82	7.50	664.0
102	1548	27.10	175.1	4.76	10.71	369.0
103	1581	23.60	201.8	4.57	6.07	362.0
104	1599	14.60	238.3	3.94	2.40	613.0
105	1623	54.30	120.5	1.77	1.66	275.0
106	1688	18.80	200.8	6.90	4.50	395.0
107	1700	18.80	198.1	5.99	4.61	326.0
108	1711	15.90	218.0	5.90	5.44	504.0
109	1731	8.90	250.0	44.80	11.30	603.0
110	1742	13.30	222.0	6.22	21.50	509.0
111	1760	15.90	259.0	3.70	10.70	621.0
112	1770	9.90	324.0	3.62	6.48	617.0
113	1784	7.90	305.0	3.76	6.51	622.0
114	1798	7.90	492.0	6.51	6.52	641.0
115	1814	8.90	625.0	9.06	5.83	609.0
116	1826	9.90	294.0	8.83	7.07	721.0
117	1842	11.00	356.0	10.40	7.59	802.0
118	1855	3.30	359.0	6.28	6.74	840.0
119	1865	4.00	355.0	10.60	7.52	874.0

Hamilton 8

Rows	Days	Flow	Acidity	Total Iron	Manganese	Sulfate
1	0	0	298	16.1	4.49	750
2	132	0	291	15	4.5	202
3	155	0	291	19	6	141
4	170	0	221	21.1	6	170
5	202	0	250	15	4.5	184
6	244	70	300	28	6	166
7	258	33	226	11	5.1	183
8	272	33	272	19	3.2	145
9	287	33	236	27	9.5	199
10	300	33	262	30	9	200
11	314	33	348	23	8	240
12	330	19	376	25	5	160
13	343	9	404	120	4.3	184
14	357	9	490	25	9	300
15	370	9	416	38	2	161
16	384	9	558	40	4.9	848
17	398	9	448	40	6.8	300
18	412	2	344	21.5	7	750
19	426	9	362	40	7.5	262
20	456	9	328	13	8.1	675
21	468	9	356	33.7	6.5	650
22	483	19	290	29	5.4	700
23	496	32	238	10.5	5.5	677
24	511	32	270	13.7	5.6	693
25	525	32	254	17.2	5.5	647
26	539	19	256	17	6.5	649
27	553	120	204	14	3.8	662
28	566	70	216	13.7	4.5	487
29	580	32	224	12.9	10	495
30	594	19	238	14.7	3.4	591
31	609	32	278	14.7	5	680
32	624	32	232	18.68	4.6	600
33	636	32	219	5.99	5.1	493
34	650	32	187	4.8	3.9	575
35	666	9	386	16	5	498
36	680	32	336	7.11	3.5	702
37	692	32	320	16.1	8.5	707
38	706	0	660	7.5	6.5	751
39	721	9	382	8.5	8	801.99
40	734	0	340	8.5	5.5	797
41	748	2	454	7.5	8.2	592.99
42	762	9	460	11.5	10.1	862.01
43	772	2	414	7.5	7	851
44	790	9	390	10	10	993
45	799	9	394	7.5	10	894.99
46	818	9	396	11	6	752
47	832	9	370	7.5	5	852

Rows	Days	Flow	Acidity	Total Iron	Manganese	Sulfate
48	846	9	430	9.8	6.2	774
49	857	9	472	9	9	819.99
50	867	9	448	8	7	817
51	874	32	270	10	6	514
52	885	9	292	8	6	543
53	899	9	372	9	8	629
54	916	49	409.4	11	5.5	449
55	930	120	324.2	8.5	3.5	365
56	944	70	282.2	6	4	363
57	958	70	283.8	10	3.8	275
58	972	32	270	4.8	2.2	322
59	989	19	112.6	11.5	6.2	390
60	1000	19	164.8	7.9	6.9	453.99
61	1015	8.9	178.4	11	7	517
62	1028	8.9	204.2	8	7	516
63	1043	8.9	247.7	10	6.3	588.99
64	1052	8.9	178.4	8.5	1.3	180
65	1070	2.4	320	8.5	12	396
66	1085	8.1	339	11	10	466
67	1098	6.7	342.1	17	9.5	791.01
68	1116	1.8	376.8	7.5	10	879
69	1126	3.3	340.9	9.5	10.5	808
70	1141	6.1	338.1	8	11	910
71	1154	5.4	291	14	13	854
72	1171	18.8	252.1	8	10	510
73	1184	72.3	151.5	7	5.3	382
74	1197	44	238.7	6	5.3	465
75	1214	18.8	254.4	8	6.5	555
76	1221	18.8	283	8	8	676.01
77	1238	9.9	309.4	8.5	10	677
78	1248	360	141.9	5.5	3.4	294
79	1266	37.1	197.9	7	4.5	436
80	1280	12.1	206.6	8.5	5	464
81	1294	46.7	198.3	9	4	334
82	1308	62.8	214.8	8.5	3.2	313
83	1322	44	202.8	8	4.3	335
84	1336	37	212	7	3.4	343
85	1351	49	189	7	4.3	345
86	1365	22	223.1	12	5.3	440
87	1379	17	232.9	7.5	5.3	413
88	1392	12	270.8	7	4.8	543
89	1407	7.9	270.8	8.3	8	621
90	1421	14.6	310.8	8.3	8.5	669.99
91	1434	25.3	221.2	8	4.5	443
92	1450	12.1	273.6	8.4	4.6	552
93	1464	11	294.1	8	6.8	618
94	1477	8.9	329	9	10	625
95	1487	7.9	232.6	7.5	11	667.01
96	1504	6.2	369.6	3.1	12.1	911

Rows	Days	Flow	Acidity	Total Iron	Manganese	Sulfate
97	1515	7.1	375.1	3.3	12.6	746
98	1526	7.1	371.6	11.6	9	827.01
99	1548	23.6	292.5	16.46	13.83	494.99
100	1581	37.1	238.2	9.15	6.02	600
101	1599	25.3	271.2	8.5	2.31	561
102	1623	140	171.2	4.56	2.18	311
103	1688	57	225.7	13.8	4.55	481.01
104	1700	44	228.3	12.22	4.7	299
105	1711	32.9	259	8.9	5.42	527
106	1731	23.6	298	0.1	8.43	616
107	1742	21.9	294	13.4	14.1	569
108	1760	21.9	329	19.7	5.51	646.99
109	1770	17.3	360	17	7.37	629
110	1784	15.9	347	16.5	7.44	634
111	1798	14.6	381	14.4	6.57	640
112	1814	14.6	394	17.1	5.6	801
113	1876	13.3	401	17.3	5.58	846.99
114	1842	12.1	401	15	7.96	858.01
115	1855	11	408	13.3	6.72	879
116	1865	8.9	451	16.2	7.82	874